

## LIVER TUMOR MARKER SEQUENCES

### CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of US Provisional Patent Application No. 60/255,674, filed December 14, 2000, which application is incorporated herein by reference as if set forth in its entirety.

### STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

**[0002]** This invention was made with US Government Support from the following agency: NIH, Grant No. CA22484. The US Government has certain rights in the invention.

### BACKGROUND OF THE INVENTION

**[0003]** Liver cancer is the fifth most common cancer worldwide. More than 400,000 cases were reported in 1990. Hepatocellular carcinoma (HCC) accounts for 80% of all liver cancer. Liver cancer can result from both viral infection and chemical exposure. Known risk factors include hepatitis B and C virus infection and exposure to aflatoxin. 1. It is not known whether distinct routes to liver cancer affect the same or different cellular pathways. No mutational model has yet been developed for liver cancer as it has been for other cancers such as colon cancer. The molecular events that precede neoplastic transformation of the liver are not well understood. With no clearly identified cause, successful treatment options are lacking. In fact, the specific genes that are deregulated in liver cancer have not yet been enumerated. This is a critical first step in developing a successful strategy for treating liver cancer.

**[0004]** There is a pressing need to understand the molecular events associated with development of liver cancer, both in humans and in animal model systems where liver cancer is extensively studied, and to provide diagnostic and therapeutic reagents for treating same.

### BRIEF SUMMARY OF THE INVENTION

**[0005]** The invention is summarized in that the applicants disclose isolated polypeptides whose expression is deregulated in liver tumor cells from human and non-human animals, relative to the expression in regenerating liver tissue, and further disclose isolated polynucleotides that encode the isolated polypeptides. As a result of this differential expression,

the polypeptides and polynucleotides are diagnostic markers for a liver cancer in humans and non-human animals. In humans, the polynucleotides map to a region of chromosome 9p.

**[0006]** In one aspect, the polypeptide is selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4.

**[0007]** In another aspect, the nucleic acid encodes a polypeptide selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4.

**[0008]** In yet another aspect, the nucleic acid has a nucleotide sequence selected from the group consisting of an intron-free coding sequence between nucleotides 35 and 859 of SEQ ID NO:1 and all of SEQ ID NO:3. The polynucleotides of SEQ ID NO:1 and SEQ ID NO:3 were obtained from murine and human genetic material, respectively. SEQ ID NO:3 is a predicted spliced cDNA sequence that has been identified in a genomic fragment of the human genome (GenBank Accession No. NT\_008335.6, which encompasses sequences previously associated with Accession No. AL391834, named in the above-mentioned provisional patent application). SEQ ID NO:3 is predicted to encode, in humans, a protein within the scope of the invention.

**[0009]** In another aspect, the polypeptide-encoding polynucleotide sequence has at least about 85% nucleotide sequence identity to the coding sequence of SEQ ID NO:1 or SEQ ID NO:3 (using the NCBI Blast 2 comparison protocol) where the polynucleotide hybridizes under stringent hybridization conditions to the polynucleotide of SEQ ID NO:1 or SEQ ID NO:3. Also within the scope of the invention is a nucleic acid having at least about 90%, and most preferably at least 95% identity to either sequence.

**[00010]** In a related aspect, a polynucleotide sequence having greater than 90% homology to the protein-encoding sequences of SEQ ID NO:1 has been identified in a region of human chromosome 9p. A putative protein encoded at that location is greater than 90% similar to the amino acid sequence of SEQ ID NO:2.

**[00011]** In another aspect, the nucleic acid hybridizes under moderately stringent hybridization conditions to a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

**[00012]** In a related aspect, the nucleic acid hybridizes under highly stringent hybridization conditions to a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

**[00013]** In another related aspect, the invention is an oligonucleotide that hybridizes under highly stringent or moderately stringent conditions to a polynucleotide of the invention.

[00014] In another aspect, a polynucleotide of the invention, whether from mice or humans or any other source, is engineered into a genetic construct downstream from a heterologous promoter not natively upstream of the polynucleotide that directs expression of the encoded protein. The genetic construct is introduced into a host cell that supports transcription of the polynucleotide and translation of the protein which can then be purified using methods known to those skilled in the art. Alternatively, the construct can be provided in an *in vitro* transcription/translation system for protein production.

[00015] In still another aspect, the invention is an antibody that specifically binds to a polypeptide of the invention.

[00016] In yet another aspect, the invention is a method for identifying modulators (inducers or suppressors) of expression of the polynucleotides and polypeptides of the invention, where the method includes the step of observing a change in level of expression of a polynucleotide or polypeptide of the invention in a host cell that expresses the polynucleotide or polypeptide after exposure of the host cell to a modulating agent.

[00017] It is an object of the present invention to provide an isolated nucleic acid and an isolated polypeptide that are associated with hepatocellular carcinoma in human and non-human animals.

[00018] In yet another aspect, the present invention provides a host cell transfected with the genetic construct described above.

[00019] In still another aspect, the invention can relate to a kit having use in a method for determining in a tumor or other cell the expression level of the polypeptide or of a nucleic acid encoding the polypeptide. The kit can contain one or more antibody directed to an epitope on the polypeptide and one or more oligonucleotide or polynucleotide that hybridizes to the nucleic acid that encodes the polypeptide. The kit can also further include additional components for use as positive or negative controls in a method for determining the expression level. Such additional components can include samples of tumor or non-tumor liver cells, or an extract of any of the foregoing, for which a level of expression of a polypeptide or a polynucleotide of the invention has been determined. Alternatively or additionally, the kit can contain a sample of one or more of a polypeptide, a polynucleotide, and an oligonucleotide of the invention for quantification purposes.

[00020] Other objects, features and advantages of the present invention will become apparent upon consideration of the following detailed description of the invention.

## BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[00021] Not applicable.

## DETAILED DESCRIPTION OF THE INVENTION

[00022] Liver cancer is generally studied in animal model systems, preferably in rodent systems, where certain strains have been bred for their high susceptibility to liver tumors. C3H/HeJ mice are highly susceptible to liver tumors after induction with diethylnitrosamine (DEN). To identify polynucleotide sequences or genes that show differential expression in liver tumor cells as compared to normal liver tissue cells, gene expression differences between liver tumors and a regenerating liver were determined using representational difference analysis (RDA: Lisitsyn, et al., Science 259:946 (1993), incorporated by reference as if set forth herein in its entirety.

[00023] In this application, the applicants report the amino acid sequences of a pair of polypeptides from murine animals and humans (and the sequences of the nucleic acids that encode the polypeptide sequences) that are highly differentially expressed in cells of human and non-human liver tumors relative to regenerating normal liver cells. The polypeptide is conveniently referred to as CRG-L1 although the designation is merely arbitrary. Further, the invention provides materials and methods for detecting expression (and changes in expression) of the nucleic acids (including mRNA, single or double stranded DNA, cDNA and the like) and production of the polypeptides, thereby facilitating use as a diagnostic marker for liver cancer and as a system for assessing putative therapeutic agents.

[00024] The polypeptides and nucleic acids of the invention can be isolated and purified from normally associated material in conventional ways such that in the purified preparation the polypeptide or nucleic acid is the predominant species in the preparation. At the very least, the degree of purification is such that the extraneous material in the preparation does not interfere with use of the polypeptide or nucleic acid of the invention in the manner disclosed herein. The polypeptide or nucleic acid is preferably at least about 85% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

[00025] Structurally, the nucleic acid sequence of murine CRG-L1 (SEQ ID NO:1) encodes a polypeptide of about 275 amino acids with a predicted molecular weight of about 30 to 35 kDA. In particular, the murine CRG-L1 includes seven putative transmembrane domains that correspond to amino acids 33-53, 62-82, 91-111, 123-143, 146-166, 174-194, and 212-232 of SEQ ID NO:2.

[00026] The nucleic acid sequences of the invention can be introduced conventionally into, and expressed in, host cells which can be prokaryotic (such as bacteria) or eukaryotic (such as yeast, insect, amphibian or mammalian cells) whereupon the transcription of nucleic acid and the properties of the encoded proteins can be assessed.

[00027] The isolation of a biologically active polypeptide that is differentially regulated in a liver tumor provides a means for assaying for inhibitors and activators (*in vivo* or *in vitro*) of such a polypeptide that can affect the development or progression of liver tumors. For example, the polypeptide can be expressed in cells and the effect of various test agents on mRNA or protein expression level relative to untreated controls can be measured. Alternatively, the level of expression can be assessed in biological samples taken directly from a human or non-human tissue.

[00028] The presence and level of such a differentially regulated protein can be readily discerned using antibodies directed to an epitope on the protein using well known methods, such as an ELISA method. The level of gene expression in a liver tumor and in regenerating liver tissue can also be measured using methods for hybridizing nucleic acids (including, without limitation, RNA, DNA, and cDNA). Such methods are generally known to those skilled in the art, but are enabled by the disclosure herein of a liver tumor-specific sequence. Because one can assess levels of expression of protein and nucleic acid, it is also, therefore, possible to develop agonists and antagonists of the encoded protein or to identify agents that affect transcription or translation of the disclosed nucleic acid sequences.

[00029] A skilled artisan understands that polypeptide sequences presented herein can vary somewhat, whether as a result, e.g., of sequencing error or allelic variation or duplication, from the sequence presented while still retaining their essential nature, that is, differential regulation in liver tumors relative to normal liver tissue. Further, the nucleic acids of the invention include conservatively modified variants of the sequences presented herein, complementary sequences, and splice variants. In view of the known degeneracy in the genetic code, the proteins disclosed can also be encoded by a large number of other polynucleotide sequences, all of which are within the scope of the invention. The polypeptides of the invention includes polymorphic variants, alleles, mutants, and interspecies homologs that (1) are differentially expressed in liver tumors, (2) bind to antibodies raised against the coding region of either disclosed polypeptide, (3) specifically hybridize under stringent hybridization conditions to a nucleic acid sequence selected from a group consisting of SEQ ID NO:1 and SEQ ID NO:3, or (4) are amplified by primers that amplify SEQ ID NO:1 and SEQ ID NO:3.

[00030] Exemplary high stringency hybridization conditions include 50% formamide, 5X SSC and 1% SDS incubated at 42 °C, or 5X SSC and 1% SDS incubated at 65 °C, followed by washing in 0.2X SSC and 0.1% SDS at 65°C. Exemplary moderate stringency hybridization conditions include 40% formamide, 1M NaCl and 1% SDS incubated at 37 °C followed by washing in 1X SSC at 45°C. These conditions are merely exemplary as one skilled in the art is readily able to discern stringent from moderately stringent hybridization conditions.

[00031] Moreover, the sequences of the invention also encompass substitutions, additions and deletions of the sequences presented where the change affects one or a few amino acids in the presented polypeptide sequences, without substantial effect upon the activity of the polypeptide.

[00032] The present invention will be better understood upon consideration of the following non-limiting example.

#### EXAMPLE

[00033] Inbred C3H/HeJ mice were bred and housed in plastic cages on corncob bedding (Bed-O'Cobs; Anderson Cob Division) and were fed Breeder Blox (Harlan). Food and acidified water were available *ad libitum*. To obtain regenerating livers, partial hepatectomies were preformed on male, six week old mice as described by Lukas, E. R., et al., Molecular Carcinogenesis 25:295-303 (1999). All papers mentioned in the example are incorporated by reference herein as if set forth in their entirety. Animals were sacrificed 36 hours after the surgery, at a time that corresponds to peak DNA synthesis, and the liver remnants were harvested.

[00034] Liver tumors were taken from male C3H/HeJ mice that had been treated with DEN (0.1 µM/g of body weight) at 12 days of age and sacrificed at 32 weeks of age.

[00035] Total RNA was extracted from liver using guanidine thiocyanate/CsCl as described by Lukas et al. PolyA mRNA was isolated from 250 µg of total RNA using Oligotex mRNA kit (Qiagen). The cDNA RDA protocol developed by Hubank, M. and D. G. Schatz, Nucleic Acid Research 22:5640-5648 (1994) was followed in detail using polyA RNA from the regenerating livers and the liver tumors. cDNA RDA is a method for cloning transcripts found in the pool of mRNA from one source, but absent from the pool of mRNA from a second source. Depending upon how the experiment is set up, one can identify novel genes that are either up-regulated or down-regulated. Briefly, mRNA is obtained from two tissues, in this case

regenerating livers and liver tumors. The cDNA RDA technique is performed on cDNA prepared using standard methods from the isolated mRNA.

**[00036]** In the first subtractive round, the representations were hybridized to each other in a 1:100 tester/driver ratio. The second and third difference products used a tester/driver ratio of 1:800 and 1:400,000 respectively. Difference products were subcloned from the second difference product subtractive round because no products were observed in the third round. Cloned products were sequenced by Big Dye Sequencing (Applied Biosystems, Inc.). Two comparisons were performed between the regenerating livers and the liver tumors. In the first comparison, the tester, provided in a more limited amount than the driver, was cDNA derived from the liver tumor. This comparison can identify genes upregulated in liver tumors. In the second comparison, the tester was cDNA derived from the regenerating liver. This comparison can identify genes upregulated in regenerating livers. A number of difference products were obtained in each comparison, indicating that some mRNAs are up-regulated in liver tumors while other mRNAs are down-regulated in liver tumors, relative to regenerating liver tissue. Many of the up-regulated and down-regulated sequences correspond to known genes. In addition, however, five novel differentially expressed transcripts were also identified.

**[00037]** The most highly differentially expressed polynucleotide sequence was examined for expression in eight mouse tissues and in four embryonic tissues using a multiple tissue cDNA panel (Clontech) according to the manufacturer's recommended protocol. The isolated RDA fragment included bases 997 through 1383 of SEQ ID NO:1. The polynucleotide was expressed most highly in heart, lung, and testes. Modest expression was seen in regenerating liver, which is consistent with the low levels observed in the RDA when compared to a liver tumor.

**[00038]** A cDNA clone of the differentially expressed polynucleotide was obtained by screening the Origene rapid-screen mouse liver cDNA library with primers designed from the isolated RDA fragment. A 4.175 kb cDNA was isolated and sequenced. At the 3' end of the cDNA, significant homology to twenty-three mouse ESTs (GenBank Accession Numbers AA048715, AA212916, AA212925, AA462019, AA462654, AA475320, AA914194, AV227941, AW490555, AW701866, AW702104, BB108761, BB627599, BB660847, BB752973, BB764116, BF144307, BF468547, BF661433, BF662488, BG109928, BG230006, and BI080821) was noted. An ATG translation initiation site was seen at base pairs 35-45 of SEQ ID NO:1, with an open reading frame extending to base 862, and followed by a stop codon. The predicted translation product is a protein of 275 amino acids having a molecular weight of

31.4 kD. Seven putative transmembrane domains were revealed using the SMART (Simple Modular Architecture Research Tool) which analyzes protein sequences for motifs. The transmembrane domains correspond to amino acids 33-53, 62-82, 91-111, 123-143, 146-166, 174-194, and 212-232 of SEQ ID NO:2. The existence of related sequences in *C. elegans* and *D. melanogaster* suggest a conserved function for the polynucleotide obtained by the inventors.

**[00039]** In the human genome data base at NCBI, clone Hs9\_8492 (Genbank Accession No. NT\_008335.6, a contig from human chromosome 9p, was shown to have areas of significant homology to the entire mouse cDNA sequence. In this clone obtained from human DNA, six exons having significant homology to the mouse polynucleotide sequence were identified. The entire sequence in humans of an open reading frame that corresponds to murine SEQ ID NO:1 is represented in clone Hs9\_8492.

**[00040]** The human sequences thus identified were joined with reference to the disclosed SEQ ID NO:1 by removing putative splice regions and pasting the remaining sequences together to join as SEQ ID NO:3 the following areas of the contig, in this order: 660483-660351, 645683-645565, 644843-644705, 634599-634459, 623266-623125, and 619093-618907 (as those sequences are numbered as of the filing date of the application in the sequence of Hs9\_8492, which contains a set of 21 as yet unlocalized pieces. This single clone includes all of the sequences that, when arranged to form coding sequence 1-825 (plus a stop codon) of SEQ ID NO:3, correspond to a polynucleotide sequence from bases 35 to 862 of SEQ ID NO:1 from mice. Further, such a sequence can encode a protein in humans that corresponds to the protein of SEQ ID NO:2. The putative human cDNA is 87% identical to the mouse sequence. If the putative human cDNA is translated, the resulting amino acid sequence is 91% similar to the corresponding portion of the mouse amino acid sequence, using the Lipman-Pearson protein alignment with a gap penalty of 4 and gap length penalty of 12.

**[00041]** Sequences corresponding to the basal promoter region have also been identified within Hs9\_8492 from 661112 to 660393. SP1 (660696-660703 and 660616-660622) and E2F (660543-660551) transcription factor binding sites have also been identified upstream of the coding sequence of the human coding sequence. Cotransfections using luciferase reporters have shown that the CRG-L1 promoter is activated by E2F1.

**[00042]** The polynucleotide and polypeptide sequences provide a skilled artisan with the ability to assess using conventional methods the expression levels of this human gene and array of tissues and more specifically to monitor the expression of the gene in human liver tumors as compared to normal human liver tissue. Likewise, antibodies directed to a portion of the human



protein can be produced and used as diagnostic agents for assessing protein levels in various human tissues including liver tumors.

**[00043]** The applicants have observed by RT-PCR analysis that, like the murine mRNA, the human CRG-L1 mRNA is upregulated relative to normal liver tissue in three different surgically-excised human hepatocellular carcinomas and one hepatocellular adenoma. The mRNA level in these samples was comparable to that observed in the HepG2 hepatocellular carcinoma cell line. Differential expression of CRG-L1 mRNA was not observed in human colon adenocarcinomas.

**[00044]** The applicants have also observed that other proteins from mouse and human libraries interact with the C-terminal domain in a yeast two-hybrid screen, namely clathrin adapter protein AP-1, megakaryocyte stimulating factor and Jab-1.

**[00045]** The present invention is not intended to be limited to the foregoing, but rather to encompass all such variations and modifications as come within the scope of the appended claims.